

1 TCGCGGGAGC CAGAGGGCCC TGCGGTCTTC GGTGGTCTTG CCAGCCCCCTC
51 CTCATCCCAG GGCCCTCCGC GCCTGTGAGG ACTCCCTCAG GTCGGCCACG
101 GGACCTGACG CAACAGGATG GACGAGTCCC CTGAGCCTCT GCAGCAGGGC
151 AGAGGGCCCG TGCCGGTCCG ACCGAGCGCG CCAGCACCCC GGGGTCTGCG
201 TGAGATGCTG AAGGCCAGGC TGTGGTGCAG CTGCTCGTGC AGTGTGCTGT
251 GCGTCCGGGC GCTGGTGCAG GACCTGCTCC CCGCCACGCG CTGGCTGCGT
301 CAGTACCGCC CGCGGGAGTA CCTGGCAGGC GACGTCATGT CTGGGCTGGT
351 CATCGGCATC ATCCTGGTGC CGCAGGCCAT CGCCTACTCA TTGCTGGCCG
401 GGCTGCAGCC CATCTACAGC CTCTATACGT CCTTCTTCGC CAACCTCATC
451 TACTTCTCTA TGGGCACCTC ACGGCATGTC TCCGTGGGCA TCTTCAGCCT
501 GCTTTGCCCTC ATGGTGGGCG AGGTGGTGGG CCGGAGGCTC CAGCTGGCCG
551 GCTTTGAGCC CTCCCAGGAC GGCCTGCAGC CCGGAGCCAA CAGCAGCACC
601 CTCAACGGCT CGGCTGCCAT GCTGGACTGC GGGCGTGA CTACGCCAT
651 CCGTGTGCGC ACCGCCCTCA CGCTGATGAC GGGGCTTTAC CAGTCTCTCA
701 TGGGCGTCTT CCGGCTGGGC TTCGTGTCCG CCTACCTCTC ACAGCCACTG
751 CTCGATGGCT TTGCCATGGG GGCTCCCGTG ACCATCCTGA CCTCGCAGCT
801 CAAACACCTG CTGGGCGTGC GGATCCCGCG GCACCAGGGG CCCGGCATGG
851 TGGTTCCTAC ATGGCTGAGC CTGCTGCGCG GCGCCGGGCA GGCCAACGTG
901 TGCGACGTGG TCACCAGCAC GGTGTGCTGC GCGGTGCTGC TAGCCGCGAA
951 GGAGCTCTCA GACCGCTACC GACACCGCCT GAGGGTGCCG CTGCCCACGG
1001 AGCTGCTGGT CATCGTGGTG GCCACACTCG TGTCGCACTT CGGGCAGCTC
1051 CACAAGCGCT TTGGCTCGAG CGTGGCTGGC GACATCCCCA CGGGTTTCAT
1101 GCGCCCTCAG GTCCAGAGC CCAGGCTGAT GCAGCGTGTG GCTTTGGATG
1151 CCGTGGCCCTT GCGCCCTGTG GCTGCCGCTT TCTCATCTC GCTGGCGGAG
1201 ATGTTCCGCC GCAGTCACGG CTACTCTGTG CGTGCCAACC AGGAGCTGCT
1251 GGCTGTGGGC TGCTGCAACG TGCTACCCGC CTTCCTCCAC TGCTTCGCCA
1301 CCAGCGCCGC CCTGGCCAAG AGCCTGGTGA AGACAGCCAC TGGCTGCCGG
1351 ACACAGCTGT CCAGCGTGGT CAGCGCCACC GTGGTGCTGC TGGTGTGCT
1401 GGCGCTGGCA CCGCTGTTC ACGACCTACA GCGAAGCGTG CTGGCCTGCG
1451 TCATCGTGGT GAGCCTGCGG GGGGCCCTGC GCAAGGTGTG GGACCTCCCG
1501 CCGCTGTGGC GGATGAGCCC GGCTGACGCG CTGGTCTGGG CAGGCACCGC
1551 GGCCACCTGT ATGCTGGTCA GCACAGAGGC CGGGCTGCTG GCTGGCGTCA
1601 TCCTCTCGCT GCTCAGCCTG GCCGGCCGCA CCCAACGCCC ACGCACCGCC
1651 CTGCTGGCCC GCATCGGGGA CACGGCCTTC TACGAGGATG CCACAGAGTT
1701 CGAGGGCCTC GTCCCTGAGC CCGCGTGCG GGTGTTCCGC TTTGGGGGGC
1751 CGCTGTACTA TGCCAAACAAG GACTTCTTCC TGCAGTCACT CTACAGCCTC
1801 ACGGGGCTGG ACGCAGGGT CATGGCTGCC AGGAGGAAGG AGGGGGGCTC
1851 AGAGACGGGG GTCCGGTGGG GAGGCCCTGC CCAGGGCGAG GACCTGGGCC
1901 CGGTTAGCAC CAGGGCTGCG CTGGTGCCCG CAGCGGCCCG CTTCACACA
1951 GTGGTCATCG ACTGCGCCCC GCTGCTGTTC CTAGACGAG CCGGTGTGAG
2001 CACGCTGAG GACCTGCGCC GAGACTACGG GGCCCTGGGC ATCAGCCTGC
2051 TGCTAGCCTG CTGCAGCCCG CCTGTGAGAG ACATTCTGAG CAGAGGAGGC
2101 TTCCTCGGGG AGGGCCCCCG GGACACGGCT GAGGAGGAG AGCTGTTCTT
2151 CAGTGTGCAC GATGCCGTGC AGACAGCAC AGCCCGCCAC AGGGAGCTGG
2201 AGGCCACCGA TGTCCATCTG TAGCAGGGCC AGGCCTGCCC AGCAGCCTCT
2251 GCTCCCTCCT GGGGACCCAC AGCAGACGTC TGCAAGCCAC TGCTGAGACC
2301 CTTCCAGGG AGGAGCCACC CAAGAGCTGC ACTCTGTGC CACAGCTGCC
2351 CTGGGGAAAC CGGGGAACCC CAACTGGGAA AGGAGGCCCT CTGATCACAC
2401 GCAGGACCCA AACACTCAGA AATCAAGAAC CTCTGCCTCC GAGACAGGCT
2451 GGCCACAGT GCTGGCTGGG CCCCAATGCA CCGTCCCTCA GCTCAGAAGG
2501 GATGGGCGCT ACCTGACGCT CAGGGTTGAC ATCTTATTTG AACAGGGTTC
2551 CCGCGCCATC ATGCAGCCTC CAAGGTGCCA AGAGGACTCC CTATGCCAG
2601 GCCTGCCCCG TGCCACCCCT GCTGGTAGGA GCCAGCGGCT CTGGCCAAGT
2651 GCACGAGGGT CTCTGTGTTT CCAGAAGGCC CCACACACCC AAGTGCCCTT
2701 CACACCTCGT GCCTCCCCCT CACAGGGTGG CCACCTGCAC CAGCGTCAGG
2751 GCGCAGGGTG CTGTGACCGA TGAGACCTCA GCTCAGCCCT CAGGTGCAGT
2801 GCGCCCTACC AGCCTGGCCA GCAGACACAC ACAGGGATGC TCACGGGTGC
2851 ACCAGGAGCC AGGTGCGGCG CAGCCAACCC TGAGCCTGCA GGGAGACCTG
2901 CAGGAAGCCC ACCGTGCCCC ATGCAGGGGC TCCCTCCAGC ACACAGCCCT
2951 CACCCAGCA CAGCCAGCAA GGACACGCTC TCCCAACAG GGTGCTTCGG
3001 CGGGAGGTGG GGGAAACAAG GGTCTTCCGA GCAGCCCCCA GCGCTCCCTT
3051 CCCATCTGTG CCTCTGTAAG GGGCTCTGGG ACGCCAGAC CCGCCCGCC
3101 GCGCACCTGG TGGTGACAAG CTCCAGCAGC CAGTGGGTCC GGACCTGCTT

FIGURE 1, page 1 of 3

3151 GATGCCGCGG TGAGGGACGG CGCCACATA GGCGAGGTTG AGCTGCTGGT
3201 CCCAGCTGAG GACGTACTGG TCAGCCTGGC TGTGTGGCAG CGGGGGGCTG
3251 GGGACAACAA AGGGGCGGCT CAGTCCCGAG CCTCAGCATG GCTGGCAGCG
3301 CGGCTGACAC ACACGTTCAA GCCCAGGACT GCCCGGGCGC AGGATCCAGG
3351 CGCTGCCCCG GCGTTCAGTG ACTAATAAAA TGACCCTTAG GGCCAGGAAA
3401 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAA (SEQ ID NO:1)

FEATURES:

5'UTR: 1-117
Start Codon: 118
Stop Codon: 2221
3'UTR: 2224

117 118 2221 2224

HOMOLOGOUS PROTEINS:**Top BLAST Hits:**

	Score	E
CRA 335001098671800 /altid=gi 11545741 /def=ref NP_071325.1 so...	1385	0.0
CRA 335001098639224 /altid=gi 11560117 /def=ref NP_071623.1 su...	1055	0.0
CRA 1000746201930 /altid=gi 6746349 /def=emb CAB69640.1 (AJ223...	654	0.0
CRA 18000004923413 /altid=gi 4557539 /def=ref NP_000103.1 sulf...	649	0.0
CRA 18000004971635 /altid=gi 627422 /def=pir A54808 diastroph...	649	0.0
CRA 154000124061898 /altid=gi 12054717 /def=emb CAC20729.1 (Y1...	647	0.0
CRA 18000005144885 /altid=gi 6015035 /def=sp O70531 DTD_RAT SUL...	631	e-179
CRA 18000004938377 /altid=gi 6681233 /def=ref NP_031911.1 dias...	622	e-177
CRA 108000024647870 /altid=gi 12730580 /def=ref XP_011158.1 so...	522	e-147
CRA 1000682322799 /altid=gi 6755022 /def=ref NP_035997.1 pendr...	416	e-115

BLAST dbEST Hits:

	Score	E
gi 10209038 /dataset=dbest /taxon=96...	1015	0.0
gi 7140527 /dataset=dbest /taxon=9606...	769	0.0
gi 5847932 /dataset=dbest /taxon=9606 ...	488	e-135

EXPRESSION INFORMATION FOR MODULATORY USE:

library source (from BLAST dbEST hits):

gi|10209038 Lung
gi|7140527 Lymph
gi|5847932 Kidney

Tissue Screening Panels:

Human heart
Human Leukocyte
Thyroid
Pituitary
Brain
Fetal brain
Adrenal gland
Testis
Kidney
Small intestine
Pancreas
Liver
Lung
Placenta
Skeletal muscle
Spleen
Hela cells

1 MDESPEPLQQ GRGPVVRRO RPAPRGLREM LKARLWCSCS CSVLCVRALV
 51 QDLLPATRWL RQYRPREYLA GDVMSGLVIG IILVPQAIAY SLLAGLQPIY
 101 SLYTSFFANL IYFLMGTSRH VSVGIFSLLC LMVGQVVDRE LQLAGFDPSQ
 151 DGLQPGANSS TLNGSAAML D CGRDCYAIRV ATALTMTGL YQVLMGVLRL
 201 GFVSAYLSQP LLDGFAMGAS VTILTSQ LKH LLGVRI PRHQ GPGMVVLTWL
 251 SLLRGAGQAN VCDVVTSTVC LAVLLAAKEL SDRYRHLRV PLPTLLVIV
 301 VATLVSHFGQ LHKRFGSSVA GDIPTGFMP QVPEPRLMQR VALDAVALAL
 351 VAAAFSISLA EMFARSHGYS VRANQELLAV GCCNVLP AFL HCFATSAALA
 401 KSLVK TATGC RTQLSSVSA TVVLLVLLAL APLFHD LQRS VLACVIVVSL
 451 RGALRKVWDL PRLWRMSPAD ALVWAGTAAT CMLVSTEAGL LAGVILSLLS
 501 LAGRTQRPR T ALLARIGDTA FYEDATEFEG LVPEPGVRVF RFGGPLYYAN
 551 KDFFLQSLYS LTGLDAGCMA ARRKEGGSET GVGE GGP AQG EDLGPVSTRA
 601 ALVPAAAGFH TVVIDCAPLL FLDAAGVSTL QDLRRDY GAL GISLLLACCS
 651 PPVRDILSRG GFLGEGPGDT AEEEQLFLSV HDAVQTARAR HRELEATDVH
 701 L (SEQ ID NO:2)

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
 N-glycosylation site

Number of matches: 2

1 158-161 NSST
 2 163-166 NGSA

[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
 Protein kinase C phosphorylation site

Number of matches: 7

1 117-119 TSR
 2 281-283 SDR
 3 370-372 SVR
 4 449-451 SLR
 5 505-507 TQR
 6 597-599 STR
 7 686-688 TAR

[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
 Casein kinase II phosphorylation site

Number of matches: 7

1 358-361 SLAE
 2 467-470 SPAD
 3 526-529 TEFE
 4 562-565 TGLD
 5 629-632 TLQD
 6 670-673 TAEE
 7 679-682 SVHD

[4] PDOC00007 PS00007 TYR_PHOSPHO_SITE
 Tyrosine kinase phosphorylation site

515-522 RIGDTAFY

[5] PDOC00008 PS00008 MYRISTYL
 N-myristoylation site

Number of matches: 15

1 76-81 GLVIGI
 2 152-157 GLQPGA
 3 156-161 GANSST
 4 218-223 GASVTI
 5 255-260 GAGQAN

6 316-321 GSSVAG
 7 476-481 GTAATC
 8 489-494 GLLAGV
 9 493-498 GVILSL
 10 563-568 GLDAGC
 11 567-572 GCMAAR
 12 576-581 GGSETG
 13 577-582 GSETGV
 14 581-586 GVGEVG
 15 660-665 GGFLGE

[6] PDOC00012 PS00012 PHOSPHOPANTETHEINE
 Phosphopantetheine attachment site

411-426 RTQLSSVVSATVLLV

[7] PDOC00870 PS01130 SULFATE_TRANSP
 Sulfate transporters signature

98-119 PIYSLYTSFFANLIYFLMGTSR

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	73	93	1.663	Certain
2	98	118	1.558	Certain
3	121	141	0.813	Putative
4	180	200	1.400	Certain
5	209	229	1.017	Certain
6	259	279	1.008	Certain
7	291	311	1.227	Certain
8	344	364	1.585	Certain
9	377	397	1.343	Certain
10	414	434	2.107	Certain
11	483	503	1.446	Certain
12	602	622	0.977	Putative
13	635	655	0.897	Putative

BLAST Alignment to Top Hit:

```
>CRA|335001098671800 /altid=gi|11545741 /def=ref|NP_071325.1| solute
  carrier family 26 (sulfate transporter), member 1 [Homo
  sapiens] /org=Homo sapiens /taxon=9606 /dataset=nraa
  /length=701
Length = 701
```

```
Score = 1385 bits (3545), Expect = 0.0
Identities = 698/701 (99%), Positives = 698/701 (99%)
Frame = +1
```

```
Query: 1      MDESPEPLQQGRGPVPVRRQRPAPRGLREMLKARLWCSCSCSVLCVRALVQDLLPATRWL 180
             MDESPEPLQQGRGPVPVRRQRPAPRGLREMLKARLWCSCSCSVLCVRALVQDLLPATRWL
Sbjct: 1      MDESPEPLQQGRGPVPVRRQRPAPRGLREMLKARLWCSCSCSVLCVRALVQDLLPATRWL 60

Query: 181    RQYRPREYLAGDVMSGLVIGIILVPQAIAYSLLAGLQPIYSLYTSFFANLIYFLMGTSRH 360
             RQYRPREYLAGDVMSGLVIGIILVPQAIAYSLLAGLQPIYSLYTSFFANLIYFLMGTSRH
Sbjct: 61     RQYRPREYLAGDVMSGLVIGIILVPQAIAYSLLAGLQPIYSLYTSFFANLIYFLMGTSRH 120

Query: 361    VSVGIFSLCLMVGQVVDRELQLAGFDPSQDGLQPGANSSTLNGSAAMLDCGRDCYAIRV 540
             VSVGIFSLCLMVGQVVDRELQLAGFDPSQDGLQPGANSSTLNGSAAMLDCGRDCYAIRV
Sbjct: 121    VSVGIFSLCLMVGQVVDRELQLAGFDPSQDGLQPGANSSTLNGSAAMLDCGRDCYAIRV 180

Query: 541    ATALTLMTGlyQVLMGVRLRGFVSAYLSQPLLDGFAMGASVTILTSQLKHLVGVRIPRHQ 720
             ATALTLMTGlyQVLMGVRLRGFVSAYLSQPLLDGFAMGASVTILTSQLKHLVGVRIPRHQ
Sbjct: 181    ATALTLMTGlyQVLMGVRLRGFVSAYLSQPLLDGFAMGASVTILTSQLKHLVGVRIPRHQ 240

Query: 721    GPGMVVLTWLSLLRGAGQANVCDVVTSTVCLAVLLAAKELSDRYRHRLRVPLPTELLVIV 900
             GPGMVVLTWLSLLRGAGQANVCDVVTSTVCLAVLLAAKELSDRYRHRLRVPLPTELLVIV
Sbjct: 241    GPGMVVLTWLSLLRGAGQANVCDVVTSTVCLAVLLAAKELSDRYRHRLRVPLPTELLVIV 300

Query: 901    VATLVSHFGQLHKRFSSVAGDIPTGFMPQVPEPRLMQRVALDAVALALVAAAFSISLA 1080
             VATLVSHFGQLHKRFSSVAGDIPTGFMPQVPEPRLMQRVALDAVALALVAAAFSISLA
Sbjct: 301    VATLVSHFGQLHKRFSSVAGDIPTGFMPQVPEPRLMQRVALDAVALALVAAAFSISLA 360

Query: 1081   EMFARSHGYSVRANQELLAVGCCNVLP AFLHCFATSAAKSLVKTATGCRTQLSSVVSA 1260
             EMFARSHGYSVRANQELLAVGCCNVLP AFLHCFATSAAKSLVKTATGCRTQLSSVVSA
Sbjct: 361    EMFARSHGYSVRANQELLAVGCCNVLP AFLHCFATSAAKSLVKTATGCRTQLSSVVSA 420

Query: 1261   TVVLLVLLALAPLFHDLQRSVLACVIVVSLRGALRKVWDL PRLWRMSPADALVWAGTAAT 1440
             TVVLLVLLALAPLFHDLQRSVLACVIVVSLRGALRKVW  PRLWRMSPADALVWAGTAAT
Sbjct: 421    TVVLLVLLALAPLFHDLQRSVLACVIVVSLRGALRKVWGFPRLWRMSPADALVWAGTAAT 480

Query: 1441   CMLVSTEAGLLAGVILSLLSLAGRTQRPRTALLARIGDTAFYEDATEFEGLVPEPGVRVF 1620
             CMLVSTEAGLLAGVILSLLSLAGRTQRPRTALLARIGDTAFYEDATEFEGLVPEPGVRVF
Sbjct: 481    CMLVSTEAGLLAGVILSLLSLAGRTQRPRTALLARIGDTAFYEDATEFEGLVPEPGVRVF 540

Query: 1621   RFGGPLYYANKDFFLQSLYSLTGLDAGCMAARRKEGGSETGVGEGGPAQGEDLGPVSTRA 1800
             RFGGPLYYANKDFFLQSLYSLTGLDAGCMAARRKEGGSETGVGEGGPAQGEDLGPVSTRA
Sbjct: 541    RFGGPLYYANKDFFLQSLYSLTGLDAGCMAARRKEGGSETGVGEGGPAQGEDLGPVSTRA 600

Query: 1801   ALVPAAAGFHTVVIDCAPLLFLDAAGVSTLQDLRRDYGALGISLLLACCSPPPVRDILSRG 1980
             ALVPAAAGFHTVVIDCAPLLFLDAAGVSTLQDLRRDYGALGISLLLACCSPPPVRDILSRG
Sbjct: 601    ALVPAAAGFHTVVIDCAPLLFLDAAGVSTLQDLRRDYGALGISLLLACCSPPPVRDILSRG 660

Query: 1981   GFLGEGPGDTAEELQLFLSVHDAVQTARARHRELEATDVHL 2103
             GFLGEGPGDTAEELQLFLSVHDAVQTARARHRELEATD HL
Sbjct: 661    GFLGEGPGDTAEELQLFLSVHDAVQTARARHRELEATDAHL 701 (SEQ ID NO:4)
```

FIGURE 2, page 3 of 4

Hammer search results (Pfam):

Model	Description	Score	E-value	N
PF00916	Sulfate transporter family	405.6	4.7e-118	1
CE00008	E00008 GUANYLIN	8.6	0.016	1
PF00497	Bacterial extracellular solute-binding prote	4.4	0.57	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF00497	1/1	338	356 ..	1	27 [.	4.4	0.57
CE00008	1/1	409	431 ..	1	24 [.	8.6	0.016
PF00916	1/1	195	505 ..	1	328 []	405.6	4.7e-118

```

1 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
51 NNNNNNNNNN NNNTGGTGAA ACCCCGTCTC TACTAAAAAT ACAAAAAATT
101 AGCCGGGCGT GGTGGCGGGT GCCTGTAGTC CCAGCTACTC GGGAGGCTGA
151 GGCAGGAGAA TCACTTGAAC CCGGGAGACA GAGCTTGACG TGAGCCGAGA
201 TCATGCCACT GTACTCCAGC CTGGGCAACA GAGCGAACT CCGTCTCAAA
251 AAAAAAAAAA TTAGCCGGGC GCGGTGGCGG GCGCCTGTAG TCCCAGCTAC
301 TCAGGAGGCT GAGGCAGGAG AATGGCGTGA ACCCAGGAGG CAGAGCTTCC
351 AGTGAGCCGA GATCACACCA CTGCATTCCG GCCTGGGTGA CAGAGCAAGA
401 CTCCGCCTCA AAAAAAAAAA AAGAAAAGGT GGGGGGCGTC TCACTATGTT
451 GACCAAGGCTG GTCTGAACT GCTGGCCTTA AGCGATCCTC CTGTCTAGGC
501 CTCCCAAGT GTTGAATTA CAGGAGTGAA CCATCGTGCC TGGCTAATAA
551 TTCCTTTTAA AAAGCAGCTT ACCCTTATTT TCACGTGTGG GCCTAATTTA
601 GTTCACTTAA AAAAATCATT TATCTTCACC CCAGCCCTAT GAGGCAGGCA
651 CTGCCGGTCC TGGTGTGTGG TAGAGGGGAG GGCAGAGGAG CCGTGAGGGT
701 GACCAGGCGC TGTGGGTGCG TGCTGGGTCC AGTCAGACCA GGACTCCTGG
751 CCAGTCACGG CACCTTGACC CCGGCAGTCC TCGCCTGGG CGGTGAGCAC
801 CACACACAGG GCTTACGCGA GCACACACG ATATGCACG ACCGGCAGCC
851 TTGGGCTGAG CCGGCTGTCA GCCTCTGCCC TGCTCCAGT TGGACCAGGC
901 TGGCTCCTTG CAGGACCAGG AGGGTGTCCG GCGACTGGAC ACGGAGACCA
951 AGCCTCCCTC AGCCCCGCT GGGTTTGAAG GCTGCTGCAC TCGACCCAG
1001 ACCCCAGAGC TGAAGGTTTA CCTGTGCTCA GCCCCTGAGC CCCCCTCC
1051 CGCTGGTCCC TAAGCCCCC CGGCAGGGCC GCAGAGCCAC AGCTGCAGCC
1101 GCTCCTGGGA GGCTGGGAGC TCCTCAGAGG CCCACACAGC TCTAACTACT
1151 ACAAGCCCTT GATTACAGTT CAACTCCCG ATCAGCCGAT CAGGTAACAT
1201 GGCTGGAGAA ACCCGTGACT CAGCAATCTG TAGGTAAATA ATTGAACACT
1251 AGAGTCCAGG GCACAGACCA CTGCCTGCAG GTTGGCGCCA CCACCCAC
1301 TCTCCCGCT GCTCGCGGA GCCAGAGGGC CCTGCGGTCC TCGTGGTCT
1351 TGCCAGCCCC TCGTCATCCC AGGGCCCTCC GCGCCTGTGA GGACTCCCTC
1401 AGGTAAGAAC CATCCTGGGC CCAGATCTCA GCTGCAGCAG AGGGGGCGT
1451 GGGAGCCGAG GCCAGAAATG CCCTGGACTC GTGGTTTCTT AGGGGCACCC
1501 TCAGGCTCAA GGCAGGTGGC CCTACTGTCC CCATTCCACA CACCTGGACC
1551 CCAGGGGCTT GGGGTGGGCT TCAGGGCATC CAGGGACCCA GTGTGGTGGG
1601 GTCTTCCAGG GAAGGGGACA CAACTCTTGC AATGTTGCCT GAGGGCCAGG
1651 ACCCCCGCTC TGTGCCCCAG GGGTGTGTG CCCAGCCTGC ATGTGTCAAC
1701 CTACCAGGCT GGGCTCACTG CCCCACACA CCGCCAGGA GACTGGAGCT
1751 CGCACACCTT GGGCCAGCGT GCAAACAGCA GGCTCAGCCC AGGCTCCAGG
1801 GTGTCTGGG CACCTGGTGT CCTGGGAGCA AAGTCTTTC CTAACGTGCG
1851 TGAGAAGAAAT GTTTAAAGTG AAAGTACATT GGAGTCTGCA AACAGGACAG
1901 ACCCGAGGCC TCACGTGGGA CCAGTCAGGC CTCTAAGCAC CGCCTCCCTA
1951 ACGCCACGGT GTTTTCCGAG ATCAAGGGAA AGGTCAGGTG CCCTTCCGGC
2001 TCTGCCGGCC CAGGGTGAAT GTGTGCAGCG GGCTGGGCCC TCTCGGTGCT
2051 GCCTCGGGAC AGTGTGTCTT GGCCGTTCCT CAGTGAGCTG GTGCAGCCTG
2101 GGAAAAAGGG CGCCTCACGT CCCAGAACTG TCTGGGCAGG GGAGACAGAC
2151 GCCATATACC CTCCTCCCCT CCCAGCTGGC CCTGATGGGG CCCCCTCCA
2201 GGCATATTCT CAGAATTCTG TCCCAAGTCC AGGCGGATGG GCTAGGCTAG
2251 TGTCTGAGTG CTGCTCCCC AGCAGACTTG GGGTCCCAGT ACCCACAAAG
2301 CTTGGCAGGG ACATAGGAGG CCTCTTCTG AGACTTCCGC CAGCCCCAGG
2351 ACCCACAGGG CAGGTGACAG AGGGGTGGGT GGAGGTGTCT CCAGGAGAGC
2401 AGGCGATGGT TTGGATGGGG GAGGGAGGGC TCTGGTGTGG GCATGGGGTG
2451 GACAGCAGGA CCGTTTGCCA ACCTGGGGAG CCAGGGAGGT GGACACGGAG
2501 CAGCTGGACT CGGCTTGCC TGCACCTGTG TCCAGTGACT GTGACATTCT
2551 GACGGTAGGC ACATGTGCGT GGTGGCAGCC CAGCCTGTTC CTGCCCCGTT
2601 GGGGAGGTGA GCTTCAGGAG GCTACAGGGT GGTTTTACG CAGGAACCGC
2651 AGAGCCAATA GGCCGGAGCT GAGCCTGGAC AGGGTGCCGC CACGCCGCC
2701 CTCAGCACTG CTGGCCTCAG CACACCCCAT GGCATGGGCT TGGTGTGTA
2751 ATCCCATCTC ACCCCACGAT GGATTCTGGA TCCAGCAGGG CCCAGCGTCC
2801 ATCCATAACG GGCAGGGGGC TGGGGCCCCG GCTGCCAGGA GAAGGCCAG
2851 CACCAATCCC CGGCCCTGGG TGGGCGAGGG GTCCGCCCCA AGGGGCCCGT
2901 TGCTGCCGGG GACCTTGTG TTTGGCCCTG GATCCGGGGG CTCCTGTGAC
2951 CATGCCCTCT TCTCGGCCGC AGGTCCGCCA CGGGACCTGA CGCAACAGGA
3001 TGGACGAGTC CCCTGAGCCT CTGCAGCAGG GCAGAGGGCC GGTGCCGGTC
3051 CGACGGCAGC GCCCAGACC CCGGGGTCTG CGTGAGATGC TGAAGGCCAG
3101 GCTGTGGTGC AGCTGCTCGT GCAGTGTGCT GTGCGTCCGG GCGTGGTGC

```

FIGURE 3, page 1 of 4

3151 AGGACCTGCT CCCC GCCACG CGCTGGCTGC GTCAGTACCG CCCGCGGGAG
3201 TACCTGGCAG GCGACGTCAT GTCTGGGCTG GTCATCGGCA TCATCCTGGT
3251 GCCGCAGGCC ATCGCCTACT CATTGCTGGC CGGGCTGCAG CCCATCTACA
3301 GCCTCTATAC GTCCCTTCTTC GCCAACCTCA TCTACTTCCT CATGGGCACC
3351 TCACGGCATG TCTCCGTGGG CATCTTCAGC CTGCTTTGCC TCATGGTGGG
3401 GCAGGTGGTG GACCGGGAGC TCCAGCTGGC CGGCTTTGAC CCCTCCAGG
3451 ACGGCTGCA GCGCGGAGCC AACAGCAGCA CCCTCAACGG CTCGGCTGCC
3501 ATGCTGGACT GCGGGCGTGA CTGCTACGCC ATCCGTGTCG CCACCGCCCT
3551 CACGCTGATG ACCGGGCTTT ACCAGGTGAG GAGCCCTGCT TGGGCACAGG
3601 GAGGGGCCCA GGGCACCCCT CCTTAGGTTT TGGCCATCCA CGAGGGCAAG
3651 GCTGGGGGCA AGCACAGGT TGGCAGAGGA GGTGCTGGCC CAAGACAGCA
3701 AGGCTTGGGC AGAGCTGGGG CGTGCCGGGG CATCCCAGGG CGAGGCACCG
3751 ACGCGGAGAG GCTGTGGATG CAGGAGGGGA GGGGCACGGG GAGCCAGTCC
3801 GGTGGGCCAT GGCCTTGCTG GGGACCAGCA GGGCAGGTGT GGCTGTGGCT
3851 CAGTGGTGCT GGACTGAGGC CATGTGGCCT CCCAGGCCCT CTGTCCTAGG
3901 TGGAGTGGGG GATGGCCTCC CCACCCCGA AGGTCTCCTG CCTTGGCCTG
3951 TCCACCTTGG CCCCCTTGG CTCCACATCT GCATGGGGGG CAGTGGGCAC
4001 CATGGGTAGG AAGCAGCAGG AAGGGGTTGC CTTCTGATAC CAGAGGTCTT
4051 AATTCTGAAA TAAACGGGC TGCTGCACGT GACAAGGGTT AGACGTGTCT
4101 ATGGCCAGCT GTGTGCACGT GTGATGCTCA CGTGGATGTC ACAGTTGTCT
4151 GCGGGCATGA GCACGCGTGG AACCAGAACT CAGGCCCGTG TGAGGAGTCT
4201 GGTTTGAAC ACACGGGGCC GCAACACAGA ATTGTCAGGT CCTGTGCCGT
4251 GACCACCACC CCTCGGGCCA TGCCAGGTGC TGGTGAAGGG CAGGTGGCTC
4301 CCGCAGGCG CCTGCTGGCC TGACCGCACT CCGTCCACAG GTCTCATGG
4351 GCGTCCCTCG GCTGGGCTTC GTGTCCGCT ACCTCTACA GCCACTGCTC
4401 GATGGCTTTG CCATGGGGGC CTCCGTGACC ATCCTGACCT CGCAGCTCAA
4451 ACACCTGCTG GCGTGCAGGA TCCCGCGGCA CCAGGGGCC GGCATGGTGG
4501 TCCTACATG GCTGAGCCTG CTGCGCGGCG CCGGGCAGGC CAACGTGTGC
4551 GACGTGGTCA CCAGCACGGT GTGCTGGCG GTGCTGCTAG CCGCGAAGGA
4601 GCTCTCAGAC CGTACCGAG ACCGCTGAG GGTGCCGCTG CCCACGGAGC
4651 TGCTGTGCAT CGTGGTGGCC AACTCGTGT CGCACTTCGG GCAGCTCCAC
4701 AAGCGCTTTG GCTCGAGCGT GGCTGGCGAC ATCCCCACGG GTTTCATGCC
4751 CCCTCAGGTC CCAGAGCCCA GGCTGATGCA GCGTGTGGCT TTGGATGCCG
4801 TGGCCCTGGC CCTCGTGGCT GCCGCTTCT CCATCTCGCT GCGGAGATG
4851 TTCGCCGCA GTCACGGCTA CTCTGTGCGT GCCAACAGG AGCTGCTGGC
4901 TGTGGGCTGC TGCAACGTGC TACCCGCTT CCTCCACTGC TTCGCCACCA
4951 GCGCCGCCCT GCGCAAGAGC CTGGTGAAGA CAGCCACTGG CTGCCGACA
5001 CAGCTGTCCA GCGTGGTCAG CGCCACCGTG GTGCTGCTGG TGCTGCTGGC
5051 GCTGGCACCG CTGTTCCACG ACCTACAGCG AAGCGTGCTG GCCTGCGTCA
5101 TCGTGGTCAG CTTGCGGGGG GCCCTGCGCA AGGTGTGGGA CCTCCGCGG
5151 CTGTGGCGGA TGAGCCCGGC TGACGCGCTG GTCTGGGCAG GCACCGCGGC
5201 CACCTGTATG CTGGTCAGCA CAGAGGCCGG GCTGCTGGCT GCGTCATCC
5251 TCTCGCTGCT CAGCCTGGCC GGCCGCACCC AACGCCACG CACCGCCCTG
5301 CTGGCCGCA TCGGGACAC GGCCTTCTAC GAGGATGCCA CAGAGTTCGA
5351 GGGCCTCGTC CCTGAGCCCG GCGTGCGGGT GTTCCGCTTT GGGGGCGCGC
5401 TGTACTATGC CAACAAGGAC TTCTTCTGCT GGTCACTCTA CAGCCTCAGC
5451 GGGCTGGACG CAGGGTGCAT GGCTGCCAGG AGGAAGGAGG GGGGCTCAGA
5501 GACGGGGGTC GGTGAGGGAG GCCCTGCCCA GGGCGAGGAC CTGGGCCCGG
5551 TTAGCACCAG GGCTGCGCTG GTGCCCAGC CGGCCGGCTT CCACACAGTG
5601 GTCATCGACT GCGCCCCGCT GCTGTTCTTA GACGCAGCTG GTGTGAGCAC
5651 GCTGCAGGAC CTGCGCCGAG ACTACGGGGC CCTGGGCATC AGCCTGCTGC
5701 TAGCCTGCTG CAGCCCGCT GTGAGAGACA TTCTGAGCAG AGGAGGCTTC
5751 CTCGGGGAGG GCGGGGGGA CACGGCTGAG GAGGAGCAGC TGTTCTCAG
5801 TGTGCAGGAT GCCGTGCAGA CAGCAGGAG CCGCCACAGG GAGCTGGAGG
5851 CCACCGATGC CCATCTGTAG CAGGGCCAGG CCTGCCAGC AGCCTCTGCT
5901 CCCTCTGGG GACCCACAGC AGACGTCTGC AAGCCACTGC TGAGACCCTT
5951 CCCAGGGAGG AGCCACCAA GAGCTGCACT CTTGTGCCAC AGCTGCCCTG
6001 GGGAAACCGG GGAACCCAA CTGGGAAAGG AGGCCCTCTG ATCACACGCA
6051 GGACCCAAAC ACTCAGAAAT CAAGAACCTC TGCCTCCGAG ACAGGCTGGC
6101 CCACAGTGCT GGCTGGGCCC CAATGCACCG TCCCTCAGCT CAGAAGGGAT
6151 GGGCCTGACC TGACGCTCAG GGTGACATC TTATTGAAC AAGGTTCCCC
6201 CGCCATCATG CAGCCTCCAA GGTGCCAAGA GGAATCCCTA TGCCAGGCC
6251 TGCCCGGTGC CCACCTGCT GGTAGGAGCC AGCGGCTCTG GCCAAGTGCA

FIGURE 3, page 2 of 4

6301 CGAGGGTCTC TGTGTTTCCA GAAGGCCCCA CACACCCAAG TGCCCTCAC
6351 ACCTCGTGCC TCCCCCTCAC AGGGTGGCCA CCTGCACCAG CGTCAGGGCC
6401 CAGGGTGCTG TGACCGATGA GACCTCAGCT CAGCCCTCAG GTGCAGTGGC
6451 CCTACCCAGC CTGGCCAGCA GACACACACA GGGATGCTCA CGGGTGCACC
6501 AGGAGCCAGG TGCGGCGCAG CCAACCCTGA GCCTGCAGGG AGACCTGCAG
6551 GAAGCCCACC GTGCCCCATG CAGGGGCTCC CTCCAGCACA CAGCCCTCAC
6601 CCCAGCACAG CCAGCAAGGA CACGCTCTCC CCAACAGGGT GCTTCGGCGG
6651 GAGGTGGGGG AACAAGGGGT CTTCGAGCA GCCCCAGCC CTCCCTCCC
6701 ATCTGTGCCT CTGTAAGGGG CTCTGGGACG CCCAGACCCT GCCCCCGCC
6751 CACCTGGTGG TGACAAGCTC CAGCAGCCAG TGGGTCCGGA CCTGCTTGAT
6801 GCCGCGGTGA GAGGCGGCGC CCACATAGGC GAGGTGAGC TGCTGGTCCC
6851 AGCTGAGGAC GTACTGGTCA GCCTGGCTGT GTGGCAGCGG GGGGCTGGGG
6901 ACAACAAAGG GGCGGCTCAG TCCCGAGCCT CAGCATGGCT GGCAGCGCGG
6951 CTGACACACA CGTTCAAGCC CAGGACTGCC CGGGCGCAGG ATCCAGGCGC
7001 TGCCCCGTGC TTCACTGACT AATAAAATGA CCCTTAGGGC CAGGAATGTG
7051 GGGAGGTCCC ATCTTCATGG GGAACGGCAG CAGCAGTAAG ACGAGGGGCC
7101 AACGCCAGCC CTGGCCCTGG CCCTGCCAGG AAGGCGGGTA CCTCAGCTCT
7151 AGGTGGAAGG AATGGGACAG GCAGGCCAGG TCCCGCTGCA GGGCCGTCCA
7201 CTCCCAGGGG AGACTCCTGG TTTACCTCAA AGAGCAGGAT CCCGGGCATC
7251 GGCTGGGCT GCAGGGGGCG GCCCAGGCTC ACGCCCCGGC GCCCACTCAG
7301 GTGGAGGACC CACCACAAA CACGGCGGGG GGCGGGCCCG GGAGAGCCAG
7351 GGCCCCAGAG GAGGGAGCTC CGGTCTCTGA AGCTCTCACA GTGCGCAGTC
7401 AGGGGGCGCC CGAGCTCTCC CCGTGGCGCC AGGGGGTCCC GGAGGCCGCG
7451 GAGCGCTCAC CAGAAGCCTG TGCTCTCCA GAAGCGCCG AGGGGCCACA
7501 GCGCGCGGGC CGCGTCCACC TGCACCAGG GCGGGGCCTC GGCCGGGGCC
7551 ACCGGGGGTG CGGCCAGGAG CGAGGCCAGG AGCGCCAGCA GCGCTGCGCG
7601 GGGGCGCAGG GGACGCATGG CCACGCGTGC TCGGGGACTG CGGGGCTTCG
7651 GGCTGCACTG CCGGTTCGCG CTCCGGGTG GAGTCTGGG GCGCACCCCA
7701 TGTGACCGCC GCCGCGGGG GGGGCTTGG TGAGGGGGCG ATGGCCGGGT
7751 GGGAGGGGTT GGGTGGCCTC GGGGAGCCTC GGGGAGCCG GAGCAGGCA
7801 GGGCTTGGAG CCCCCTTCC TTGCGGGCCT CAGGGGTGC TGTGAGGACC
7851 GATGACTCGG AAAGCGCTCA GAAGAAGCT TCGCCGTTG GTGCTATGTG
7901 AGTTGAGCCA TTAAGTCTT GTTTTCTCT GTTTTGTGT GTTTTGTAGA
7951 CAGAGTCTTG CTTTGTGCG CAGGCTGAGG TGCAGTGGC CGATCTCAGC
8001 TCACTGCAAC CTCCATCTCC GGGGCTTCAG CGATTTTCTC ACCCCAGCCT
8051 CCTGAGTAAA GCGTGCGCTT TAGCAGGAAG GAGAATTACC CCAGAAGAGC
8101 AACTGGGCC CTCCTTACAC TTGGCTTCAG ATCCATGGAT TCAACCAAGC
8151 AGACTGAAAA TATTGTTTAA AAGCCAAAGC AATACGAAAT AATACATATT
8201 TTAACAAAT ACAGTATAAC AGCTATTTAC AGAGCATTTA CATTGTTTTA
8251 GGGACTATAA GTAATCTTGA TTTAACTAC ACAGTAGGAT GTGCGTAGGT
8301 AATGTGCAAA TACTGTGCCA TTTTATATCA AGTACTTGAG CACCTGCAAA
8351 TTTTGGTATC TGGGAGGGTC CTGGAACCAA TACCCGAGG ATACCATGGG
8401 ACAACTGTAG TACATGTGTA GTCCATGTAT GCATGTGTGA ATCCAAGCAA
8451 ACATTGTATA AAAATAATAA TGGAAAGAAC AGGCTTGGT CGGTGGCTCA
8501 CACCTGCAAT CCCAGCACTT TGGAATTGCA GGCCAACACG GGAGGATCAC
8551 TTGAGGCTG GAGTTTAAA TCGGCCTGGG AGATGTACCA AGACCCATC
8601 TGTACAAAAA AAAAATTTAG CCAGATGCGA TGGTATATGC CTGTGAGGCC
8651 CAGCTACCCA CGAAATTGAG GTGGGAGATT GCTTGAGCTT AGGAGTTCAA
8701 GGCTGAGACG GGCCATGATC ACACCACTAC ATTCCAGCCT GGTGACAAA
8751 ATGAGACCCC ATCTCTAAAA AAAGAAAAGA AAAAAAGAAC AGTCTACTAA
8801 CAAAACGAAA ATACTGGACA ATAATCCTCT CTAAGTTGGG AGAAGGATAA
8851 TTAGAGTTAC AGTGTCT (SEQ ID NO:3)

FEATURES:

Start: 3000
Exon: 3000-3575
Intron: 3576-4340
Exon: 4341-5867
Stop: 5868

One non-coding Exon in the 5' UTR:
(query = cDNA sequence; subject = genomic sequence)
Score = 174 bits (88), Expect = 7e-46
Identities = 91/92 (98%)
Strand = Plus/Plus

Query: 1 tcgcgggagccagagggccctgcggtcctcggtggtcttgccagcccctcctcatcccag 60
|||||
Sbjct: 1313 tcgcgggagccagagggccctgcggtcctcggtggtcttgccagcccctcgtcatcccag 1372

Query: 61 ggccctccgcgcctgtgaggactccctcaggt 92
|||||
Sbjct: 1373 ggccctccgcgcctgtgaggactccctcaggt 1404

CHROMOSOME MAP POSITION:
Chromosome 4

ALLELIC VARIANTS:
C/G nucleotide polymorphism at genomic position 1363 (in non-coding exon; see
cDNA/genomic sequence alignment above for the non-coding exon)

V/A amino acid polymorphism at protein position 699